🚀 Getting Started

Contents

Github oumi   Blog oumi   Discord 107 online   pypi package 0.1.3   License Apache 2.0   🐙 Stars 6.9k

code style black   Ⓟ pre-commit enabled   About oumi
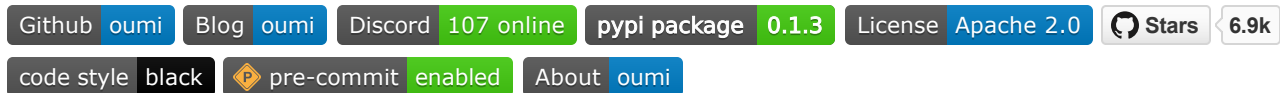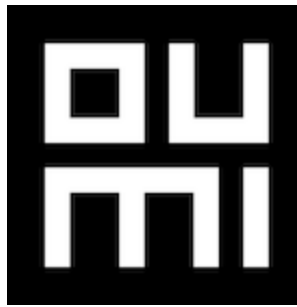
Everything you need to build state-of-the-art foundation models, end-to-end.

1 GITHUB TRENDING
#1 Repository Of The Day

Oumi is a fully open-source platform that streamlines the entire lifecycle of foundation models - from data preparation and training to evaluation and deployment. Whether you're developing on a laptop, launching large scale experiments on a cluster, or deploying models in production, Oumi provides the tools and workflows you need.

With Oumi, you can:

- 🚀 Train and fine-tune models from 10M to 405B parameters using state-of-the-art techniques (SFT, LoRA, QLoRA, DPO, and more)
- 🤖 Work with both text and multimodal models (Llama, DeepSeek, Qwen, Phi, and others)
- 🔄 Synthesize and curate training data with LLM judges
- ⚡ Deploy models efficiently with popular inference engines (vLLM, SGLang)
- 📊 Evaluate models comprehensively across standard benchmarks
- 🌍 Run anywhere - from laptops to clusters to clouds (AWS, Azure, GCP, Lambda, and more)
- 🔌 Integrate with both open models and commercial APIs (OpenAI, Anthropic, Vertex AI, Parasail, ...)

All with one consistent API, production-grade reliability, and all the flexibility you need for research. Oumi is currently in <u>beta</u> and under active development.

Notebook	Try in Colab	Goal
🎯 **Getting Started: A Tour**	Open in Colab	Quick tour of core features: training, evaluation, inference, and job management
🔧 **Model Finetuning Guide**	Open in Colab	End-to-end guide to LoRA tuning with data prep, training, and evaluation
📚 **Model Distillation**	Open in Colab	Guide to distilling large models into smaller, efficient ones
📋 **Model Evaluation**	Open in Colab	Comprehensive model evaluation using Oumi's evaluation framework
☁️ **Remote Training**	Open in Colab	Launch and monitor training jobs on cloud (AWS, Azure, GCP, Lambda, etc.) platforms
📈 **LLM-as-a-Judge**	Open in Colab	Filter and curate training data with built-in judges
🔄 **vLLM Inference Engine**	Open in Colab	Fast inference at scale with the vLLM engine

If you need a comprehensive platform for training, evaluating, or deploying models, Oumi is a great choice.

Here are some of the key features that make Oumi stand out:

- 🔧 **Zero Boilerplate**: Get started in minutes with ready-to-use recipes for popular models and workflows. No need to write training loops or data pipelines.
- 🏢 **Enterprise-Grade**: Built and validated by teams training models at scale
- 🎯 **Research Ready**: Perfect for ML research with easily reproducible experiments, and flexible interfaces for customizing each component.
- 🌐 **Broad Model Support**: Works with most popular model architectures - from tiny models to the largest ones, text-only to multimodal.
- 🚀 **SOTA Performance**: Native support for distributed training techniques (FSDP, DDP) and optimized inference engines (vLLM, SGLang).
- 🤝 **Community First**: 100% open source with an active community. No vendor lock-in, no strings attached.

While you can dive directly into any section that interests you, we recommend following the suggested path below to get the most out of Oumi.

Category	Description	Links
🚀 Getting Started	Get up and running quickly with Oumi	→ Quickstart → Installation → Core Concepts
📚 User Guides	Learn how to use Oumi effectively	→ Training → Inference → Evaluation
🤖 Models	Explore available models and recipes	→ Overview → Recipes → Custom Models
🔧 Development	Contribute to Oumi	→ Dev Setup → Contributing → Style Guide
📖 API Reference	Documentation of all modules	→ Python API → CLI

Oumi is a community-first effort. Whether you are a developer, a researcher, or a non-technical user, all contributions are very welcome!

- To contribute to the `oumi` repository, please check the `CONTRIBUTING.md` for guidance on how to contribute to send your first Pull Request.
- Make sure to join our Discord community to get help, share your experiences, and contribute to the project!
- If you are interested by joining one of the community's open-science efforts, check out our open collaboration page.

If you encounter any issues or have questions, please don't hesitate to:

1. Check our FAQ section for common questions and answers.
2. Open an issue on our GitHub Issues page for bug reports or feature requests.
3. Join our Discord community to chat with the team and other users.